

(11)特許出願公開番号

特開平7-104780

(43)公開日 平成7年(1995)4月21日

(51) Int.Cl.⁸

G 1 0 L 3/00

識別記号

5 3 1 J 9379-5H

Z 9379-5H

5 6 1 A 9379-5H

庁内整理番号

FI

技術表示箇所

審査請求 有 請求項の数5 O L (全 8 頁)

(21)出願番号

特願平5-247835

(22) 出願目

平成5年(1993)10月4日

(71)出願人 593118597

株式会社エイ・ティ・アール音声翻訳通信
研究所

京都府相楽郡精華町大字乾谷小字三平谷 5
番地

(72)発明者 山口 耕市

京都府相楽郡精華町大字乾谷小字三平谷5番地 株式会社エィ・ティ・アール音声翻訳通信研究所内

(72)發明者 嵯峨山 茂樹

東京都保谷市中町5丁目5番10号

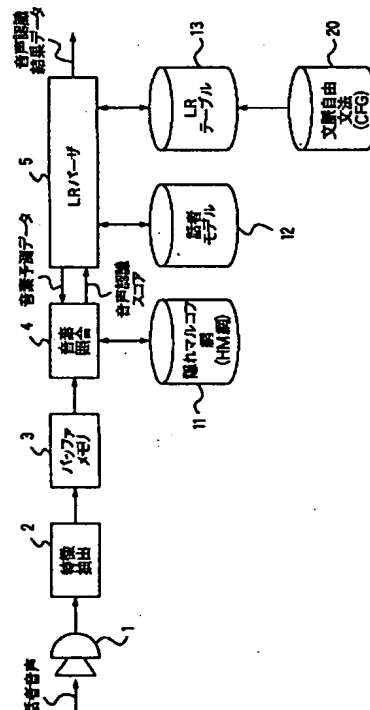
(74) 代理人 弁理士 青山 葆 (外2名)

(54) 【発明の名称】 不特定話者連続音声認識方法

(57) 【要約】

【目的】 従来例に比較して計算量を軽減することができ、しかも音声認識率を大幅に改善することができる不特定話者連続音声認識方法を提供する。

【構成】 複数 m 人の話者に対応して複数 m 個の発声内容の仮説が存在し、その後各仮説は話者毎にそれぞれ時間経過につれて成長させた音素解析木を用いて、入力された1人の話者の発声内容に基づいて発声内容と話者の2方向を同時にサーチの対象としてビームサーチしながら音声認識を連続的に実行し、音声認識動作中に計算された尤度が所定のしきい値以上のときはこれ以降の尤度計算を行い認識候補として残す一方、それ以外のときは、尤度がしきい値未満となった枝に接続される1つ又は複数の枝を枝刈りしてこれ以降認識候補として残さないで尤度の計算を行わず、音素解析木の終端部において最大の尤度又は所定値以上の尤度を有する発声内容と話者とを同時に音声認識結果として決定する。



【特許請求の範囲】

【請求項1】 不特定話者の音声を連続的に認識する不特定話者連続音声認識方法であって、

複数 m 人の話者に対応して複数 m 個の発声内容の仮説が存在し、その後各仮説は話者毎にそれぞれ時間経過につれて成長させた音素解析木を用いて、入力された1人の話者の発声内容に基づいて発声内容と話者の2方向を同時にサーチの対象としてビームサーチしながら音声認識を連続的に実行し、

上記音声認識動作中に計算された尤度が所定のしきい値以上のときはこれ以降の尤度計算を行い認識候補として残す一方、所定のしきい値未満となったときは、尤度がしきい値未満となった枝に接続される1つ又は複数の枝を枝刈りしてこれ以降認識候補として残さないで尤度の計算を行わず、上記音素解析木の終端部において最大の尤度又は所定値以上の尤度を有する発声内容と話者とを同時に音声認識結果として決定することを特徴とする不特定話者連続音声認識方法。

【請求項2】 上記音声認識動作中又は完了後は、最大の尤度を有する話者を選出し、選出した話者を標準話者として話者モデルに対して話者適応することを特徴とする請求項1記載の不特定話者連続音声認識方法。

【請求項3】 上記音声認識動作中又は完了後は、所定の上位複数個の尤度を有する話者を選出し、選出した話者群を標準話者群として話者モデルに対して話者適応することを特徴とする請求項1記載の不特定話者連続音声認識方法。

【請求項4】 上記音声認識動作完了後に、最大の尤度を有する1人の話者の発声内容を選出し、選出した発声内容を教師信号として用いて話者モデルに対して話者適応することを特徴とする請求項1記載の不特定話者連続音声認識方法。

【請求項5】 請求項1記載の不特定話者連続音声認識方法において、

上記音声認識動作完了後に、最大の尤度を有する1人の話者の発声内容を選出し、選出した1人の話者を入力話者として判断することによって話者識別することとする話者識別方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、不特定話者の音声を連続的に認識する不特定話者連続音声認識方法に関する。

【0002】

【従来の技術及び発明が解決しようとする課題】従来の不特定話者音声認識システム（以下、第1の従来例という。）が、例えば、Madeleine Bates et al.: "Design and Performance of HARC, The BBN Spoken Language Understanding System", Proc. ICSLP-92, pp. 241-244 (1992年)に開示されている。この第1の従来例においては、「不特定話者音響モデル」が用いられ、多数の話者

の音声データを混合してしばしば学習に用いることが多いために、広がり過ぎた混合分布によって認識性能が劣化する恐れがある。また、ユーザは音声の登録なしに使えるという利点がある反面、認識しにくい話者に対して性能を改善することができないという欠点がある。

【0003】また、英語の母音認識を行う方法（以下、第2の従来例という。）が、P. Niyogi and V. W. Zue:

"Correlation Analysis of Vowels and their Application to Speech Recognition", Proc. Eurospeech-91, p. 1253-1256 (1991年)に開示されている。この第2の従来例においては、母音の相関的な解析を音素認識に適用しているが、文法を用いて連続音声認識を行っていない。

【0004】さらに、入力された話者音声に対して、1つの男子音響モデルと1つの女子音響モデルとを用いて音声認識処理を並列に実行し、入力音声の最後において最高の音声認識スコアを有した認識候補を認識結果とする方法（以下、第3の従来例という。）が、例えば、V. Abrash et al., "Connectionist Gender Adaptation in a Hybrid Neural Network/Hidden Markov Model Speech Recognition System", Proc. ICSLP-92, pp. 911-914 (1992年)において開示されている。この第3の従来例においては、音響モデルを2種類しか用いていないために、不特定話者音響モデルのような広がり過ぎた音響モデルによる認識性能の劣化が考えられる。また、複数の音響モデルを用いて音声認識処理を並列に実行する方法（以下、話者並列方法という。）を用いているために計算量が大きくなり、認識効率が比較的に悪いという問題点があった。

【0005】上記話者並列方法において、標準パターンを話者 S_i 毎に設定してそれぞれ音声認識とビームサーチとを並列に実行させ、入力音声の最後に最も高いスコアの仮説を認識結果とすることが考えられる。図3にこの話者並列方法を用いた従来の不特定話者認識モードにおける音素解析木を示す。図中、各々の枝に沿って記されたアルファベットは予測・照合された音素を表す。図3の音素解析木を用いて音声認識を実行する装置においては、各枝毎に複数人分の話者の尤度を予め記憶しており、複数人分のモデルに対して最尤度を計算するためには、最後の音声までの音素解析木の最右側の端部まで並列にすべての複数人分のモデルに対する計算を実行する必要があるため、計算量が大きくなり、認識効率が悪くなるという問題点があった。

【0006】本発明の目的は以上の問題点を解決し、従来例に比較して計算量を軽減することができ、しかも音声認識率を大幅に改善することができる不特定話者連続音声認識方法を提供することにある。

【0007】

【課題を解決するための手段】本発明者は、入力される発声音声は不特定話者の発声音声であっても、その話者

3

は発声を通じて同一であることに着目して以下に示す本発明を説明した。本発明に係る請求項1記載の不特定話者連続音声認識方法は、不特定話者の音声を連続的に認識する不特定話者連続音声認識方法であって、複数m人の話者に対応して複数m個の発声内容の仮説が存在し、その後各仮説は話者毎にそれぞれ時間経過につれて成長させた音素解析木を用いて、入力された1人の話者の発声内容に基づいて発声内容と話者の2方向を同時にサーチの対象としてビームサーチしながら音声認識を連続的に実行し、上記音声認識動作中に計算された尤度が所定のしきい値以上のときはこれ以降の尤度計算を行い認識候補として残す一方、所定のしきい値未満となったときは、尤度がしきい値未満となった枝に接続される1つ又は複数の枝を枝刈りしてこれ以降認識候補として残さないで尤度の計算を行わず、上記音素解析木の終端部において最大の尤度又は所定値以上の尤度を有する発声内容と話者とを同時に音声認識結果として決定することを特徴とする。

【0008】また、請求項2記載の不特定話者連続音声認識方法は、請求項1記載の不特定話者連続音声認識方法において、上記音声認識動作中又は完了後は、最大の尤度を有する話者を選出し、選出した話者を標準話者として話者モデルに対して話者適応することとする。さらに、請求項3記載の不特定話者連続音声認識方法は、請求項1記載の不特定話者連続音声認識方法において、上記音声認識動作中又は完了後は、所定の上位複数個の尤度を有する話者を選出し、選出した話者群を標準話者群として話者モデルに対して話者適応することとする。また、請求項4記載の不特定話者連続音声認識方法は、請求項1記載の不特定話者連続音声認識方法において、上記音声認識動作完了後に、最大の尤度を有する1人の話者の発声内容を選出し、選出した発声内容を教師信号として用いて話者モデルに対して話者適応することとする。さらに、請求項5記載の話者識別方法は、請求項1記載の不特定話者連続音声認識方法、上記音声認識動作完了後に、最大の尤度を有する1人の話者の発声内容を選出し、選出した1人の話者を入力話者として判断することによって話者識別することとする。

【0009】

【実施例】以下、図面を参照して本発明に係る実施例について説明する。本実施例の不特定話者連続音声認識方法は、図2にその一例を示す音素解析木上のビームサーチする方法を用いて、発声内容と話者の2方向をサーチの対象とし、尤度が最大である発声内容と話者とを同時

$$P(w, S_i | y) = P(y | w, S_i) P(w) P(S_i) / P(y)$$

ここで、 $P(S_i)$ は入力話者が第i番目の話者である先験確率である。本実施例において、確率 $P(w)$ は、好ましくは0.1に設定する。また、不特定話者音声認識タスクを対象としているので、 $P(S_i)$ はすべての

4

に決定して音声認識することを特徴とする。音声認識動作中に計算された尤度が所定のしきい値未満となったときは、尤度がしきい値未満となった枝に接続される図2の右方向の1つ又は複数の枝を枝刈りする。そして、上記音素解析木の終端部において最大の尤度又は所定値以上の尤度を有する発声内容と話者とを同時に音声認識結果として決定する。

【0010】本実施例の不特定話者連続音声認識方法について説明するために、まず、定式化を行う。1つの発話が多数の話者の声で構成されることは現実にはほとんどあり得ない。すなわち、音声認識システムの対象が不特定話者であっても、1つの文や単語列などの中ではすべての音素は同一の話者によって発声されるという制約がある。この原理的な制約を「話者一貫性原理」と呼ぶ。

【0011】まず、話者一貫性原理の数学的定式化を行なう。 w を単語列 $w = w_1, w_2, \dots, w_n$ とおく。音響パラメータの時系列 y が与えられたならば、音声認識処理は次の数1を満足する最大の尤度を有する単語列（最尤単語列） w_a を見付けることである。ここで、「 w_a 」の「 a 」は最大尤度を示す添字である。

【数1】

$$P(w_a | y) = \max_w P(w | y)$$

ここで、右辺の \max は単語列 w に関して確率 $P(w | y)$ の最大のものを表す。 $P(w | y)$ は音響パラメータの時系列 y が与えられたときに単語列 w が見つかる確率であり、 $P(w_a | y)$ は、単語列 w に関する確率 $P(w | y)$ の中で最大(\max)の確率を有する最尤単語列 w_a の確率である。

【0012】ここで、1つの文や単語列などの中ではすべての音素は同一の話者によって発声されるという上記話者一貫性原理を、数1に適用すると次の数2を得る。

【数2】

$$P(w_a, S_i | y) = \max_{w, S_i} P(w, S_i | y)$$

【0013】ここで、右辺の \max は単語列 w とi番目の話者 S_i に関する確率 $P(w, S_i | y)$ の最大のものを表す。 S_i は第i番目の話者(i は1, 2, ..., mのいずれかである。)を表す。上記数2の右辺をベイズの定理を用いて書き換えることによって次の数3を得る。

【数3】

話者($i = 1, 2, \dots, m$)に対して等確率とする。

【0014】上記数2と数3から、音声認識処理の目的は $P(w) P(S_i) P(y | w, S_i)$ を最大にする単語列 w_a および話者 S_a を見付けることに相当するの

で、次の数4を得る。

$$P(wa)P(Sa)P(y|wa, Sa) = \max_{w, S_i} \{P(w)P(S_i)P(y|w, S_i)\}$$

【0015】ここで、右辺のmaxは、単語列wとi番目の話者 S_i に関する $\{P(w)P(S_i)P(y|w, S_i)\}$ の最大のもを表わす。 $P(y|w, S_i)$ は単語列wがある話者(i番目の話者) S_i によって制限されていることを意味する。すなわち、本方法は発話内容の単語列wに加え、話者 $\{S_i\}$ も探索の対象とする。認識動作完了とともに、選出された最大の尤度を有する

【数4】

話者が、以後の話者適応で使うのに適した標準話者 S_a として選出される。

【0016】上記話者一貫性原理の別の定式化も可能であって、最終的な尤度はすべての話者を対象とすることによって、次の数5を得る。

【数5】

$$P(wa)P(y|wa) = \max_{i=1}^m \{P(w)\sum P(S_i)P(y|w, S_i)\}$$

【0017】ここで、右辺の Σ は $i=1$ から m までの代数和である。数5は、すべての話者による寄与を考慮に入れているということを意味する。数5の場合においては、ビームサーチのときに枝刈りされて出て来ないパスが出てくるので、最大の尤度の単語列のみならず、別の話者方向を加えて別のパスも加算してサーチする。この場合、最大の尤度を有する話者 S_m は別途求める必要がある。

【0018】以上に述べた話者バージング方法を、図1に示すSSS(Successive State Splitting: 逐次状態分割法) - LR(left-to-right rightmost型)不特定話者連続音声認識装置に適用する。この装置は、メモリ11に格納された隠れマルコフ網(以下、HM網という。)と呼ばれる音素環境依存型の効率のよいHMMの表現形式を用いている。また、上記SSSにおいては、音素の特徴空間上に割り当てられた確率的定常信号源(状態)の間の確率的な遷移により音声パラメータの時間的な推移を表現した確率モデルに対して、尤度最大化の基準に基づいて個々の状態をコンテキスト方向又は時間方向へ分割するという操作を繰り返すことによって、モデルの精密化を逐次的に実行する。

【0019】図1において、話者の発声音声はマイクロホン1に入力されて音声信号に変換された後、特徴抽出部2に入力される。特徴抽出部2は、入力された音声信号をA/D変換した後、例えばLPC分析を実行し、対数パワー、16次ケプストラム係数、 Δ 対数パワー及び16次 Δ ケプストラム係数を含む34次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ3を介して音素照合部4に入力される。

【0020】音素照合部4に接続されるHM網メモリ11内のHM網は、各状態をノードとする複数のネットワークとして表され、各状態はそれぞれ以下の情報を有する。

- (a) 状態番号
- (b) 受理可能なコンテキストクラス
- (c) 先行状態、及び後続状態のリスト
- (d) 出力確率密度分布のパラメータ

(e) 自己遷移確率及び後続状態への遷移確率

【0021】なお、本実施例において、話者バージングのためのHM網は、各分布がどの話者に由来するかを特定する必要があるため、所定の話者混合HM網を変換して作成する。ここで、出力確率密度関数は34次元の対角共分散行列をもつ混合ガウス分布であり、各分布はある特定の話者のサンプルを用いて学習されている。

【0022】音素照合部4は、音素コンテキスト依存型LRパーザ(以下、LRパーザという。)5からの音素照合要求に応じて音素照合処理を実行する。このときに、LRパーザ5からは、音素照合区間及び照合対象音素とその前後の音素から成る音素コンテキスト情報が渡される。音素照合部4は、受け取った音素コンテキスト情報に基づいてそのようなコンテキストを受理することができるHM網上の状態を、先行状態リストと後続状態リストの制約内で連結することによって、1つのモデルが選択される。そして、このモデルを用いて音素照合区間内のデータに対する尤度が計算され、この尤度の値が音素照合スコアとしてLRパーザ5に返される。このときに用いられるモデルは、隠れマルコフモデル(以下、HMMという。)と等価であるために、尤度の計算には通常のHMMで用いられている前向きパスアルゴリズムをそのまま使用する。

【0023】文脈自由文法データベースメモリ20内の所定の文脈自由文法(CFG)を公知の通り自動的に変換してLRテーブルを作成してLRテーブルメモリ13に格納される。LRパーザ5は、例えば音素継続時間長モデルを含む話者モデルメモリ12と上記LRテーブルを参照して、入力された音素予測データについて左から右方向に、後戻りなしに処理する。構文的にあいまいさがある場合は、スタックを分割してすべての候補の解析が平行して処理される。LRパーザ5は、LRテーブルメモリ13内のLRテーブルから次にくる音素を予測して音素予測データを音素照合部4に出力する。これに応答して、音素照合部4は、その音素に対応するHM網メモリ11内の情報を参照して照合し、その尤度を音声認識スコアとしてLRパーザ5に戻し、順次音素を接続

していくことにより、連続音声の認識を行っている。複数の音素が予測された場合は、これらすべての存在をチェックし、ビームサーチの方法により、部分的な音声認識の尤度の高い部分木を残すという枝刈りを行って高速処理を実現する。入力された話者音声の最後まで処理した後、詳細後述するように、全体の尤度が最大のもの又は所定の上位複数個のものを認識結果データ又は結果候補データとして出力する。

【0024】本実施例の連続音声認識装置においては、音素解析木上のビームサーチを採用している。図2は話者パーキング認識モードにおける音素解析木を示し、各々の枝に沿って記されたアルファベットは予測・照合された音素を表わす。ビームサーチによってある話者の仮説がすべて枝刈りされてしまうことがあるため、数5は近似的にしか用いることができない。従って、本実施例の実際の装置では数4を採用する。まず最初に複数 m 人の話者 S_i ($i=1, 2, \dots, m$) に対応して m 個の仮説が存在する。その後、各仮説は話者毎にそれぞれ音素に同期して成長し、ビームサーチにより枝刈りされる。すなわち、認識動作中に計算された尤度が所定のしきい値以上のときはこれ以降の尤度計算を行い認識候補として残すが、一方、所定のしきい値未満となったときは、尤度がしきい値未満となった枝に接続される図2の右方向の1つ又は複数の枝を枝刈りして、これ以降、認識候補として残さず、尤度の計算を行わない。そして、上記音素解析木の終端部において最大の尤度又は所定値以上の尤度を有する発声内容と話者を同時に音声認識結果として決定する。本実施例においては、音素解析木は音素に同期して成長されているが、これに限らず、時間軸のフレームに同期して成長させてもよい。

【0025】従って、本実施例においては、発声内容と話者の2方向を同時にサーチの対象とし、最大の尤度の発声内容と話者を同時に決定することを特徴とし、上述のように、ビームサーチによって発声内容と話者の2方向の仮説を枝刈りする。

【0026】また、音声認識動作中又は完了後は、最大の尤度の話者を選出し、選出された話者を標準話者として話者適応してもよい。話者適応は、具体的には、入力話者の発声音声の少量の音声データを用いて以下のように行われる。標準話者の特徴ベクトルを入力話者の特徴ベクトル空間へ写像する移動ベクトルをHMMの学習により求め、この写像を用いて話者適応を行う。この方法は、この写像の連続性と滑らかさを仮定することにより、少量の音声データによる話者適応の高精度化を実現している。すなわち、学習により移動ベクトルが得られなかった特徴ベクトルについては、近傍の特徴ベクトルをもちいて内挿する。また、データ不足に対しては、得られた移動ベクトルに平滑化を施す。さらに、尤度が所定の上位複数個の話者を選出し、音声認識動作中又は完了後は、選出された話者群を標準話者群として話者適応

してもよい。

【0027】本実施例の方法は、認識結果の最尤単語列 wa を利用することができる点から、本方法は、言語制約を取り入れた教師信号なしの話者適応装置に適用することができる。また、最大の尤度の1人の話者の発声内容を選出し、認識動作完了後に、選出された発声内容を教師信号として用いて話者適応してもよい。さらに、最大の尤度の1人の話者の発声内容を選出し、認識動作完了後に、選出された1人の話者を入力話者として判断することによって話者識別してもよく、これにより話者識別装置を構成してもよい。

【0028】本発明者は、本実施例の図1に示す装置を用いて文節単位でシミュレーションを行い、本発明に係る話者パーキング方法と、従来技術の不特定話者法、及び第3の従来例の話者並列方法との認識性能を比較した。

【0029】まず、当該シミュレーションの条件は以下の通りである。評価話者は12名（男性5名、女性7名）であり、評価データは345文節からなる「国際会議予約タスク」を用いた。従って、全データ数は4, 140文節となる。文脈自由文法のルール数は2, 813であり、音素パープレキシティは3. 3であった。HM網の状態数は200であり、混合数は20であった。ビーム幅は最大1, 200に設定した。なお、話者並列方法ではビーム幅2, 400でも行なっており、このときの1話者あたりのビーム幅はそれぞれ60と120に相当する。

【0030】次いで、HM網は以下のようにして作成した。まず初期モデルとして170名（男性85名、女性85名）分の特定話者HM網を作成した。次に、この170名分のHM網からクラスタリングによって20個のHM網（男性11名、女性9名）を選出した。最後に各クラスに属するメンバーの話者のサンプルを用いて、VFS法によって再学習することで話者クラスHM網を作成し、それらを話者混合して不特定話者HM網とした。

【0031】上記3つの方法による認識結果を表1に示す。この結果では、不特定話者を、話者パーキング法はわずかに上回った程度である。本発明で用いた話者一貫性原理は、対象とする話者の種類が多い場合に有効であると考えられ、また話者パーキング方法を用いた方が有意に上回っている評価話者も存在することから、今後、本方式の本質的な有効性が明らかになると期待できる。一方、話者並列方法は話者毎の仮説に対してビームサーチを行なうため、枝刈りの効率が悪く、無駄な話者の仮説が生き残っていることが多い。従って、ビーム幅を2, 400に設定してもなお、本発明の話者パーキング方法と従来技術の不特定話者に及ばない。

【0032】

【表1】

文節認識率 (%)

方法	不特定話者	話者パージング	話者並列	話者並列
ビーム幅	1 2 0 0	1 2 0 0	1 2 0 0	2 4 0 0
1 位	8 3 . 3	8 3 . 9	6 2 . 3	7 6 . 1
1 位～5 位	9 4 . 9	9 5 . 2	7 0 . 5	8 6 . 1

【0033】上記シミュレーションにおいては、文節単位で音声認識を行っているが、これに限らず、文単位又は複数の文単位で音声認識を行ってもよい。

【0034】以上説明したように、不特定話者の発声であっても、話者は発声を通して同一である点に着眼した不特定話者連続音声認識方法である「話者パージング」方法を発明した。本発明者による上記SSS-LR連続音声認識装置上で不特定話者音声認識シミュレーションを行ない、従来技術の不特定話者法との認識性能を比較した。今回のような小規模の実験においては不特定話者法の認識率をわずかに上回った程度であったが、本発明の方法は、将来、対象とする話者のバラエティが広い場合にその効果を発揮し、音声認識率を大幅に改善することができると考えられる。

【0035】本発明に係る本実施例の不特定話者連続音声認識方法は、以下の特有の利点を有する。

(a) 順位の低い仮説しかもたない話者は枝刈りされ、その時点から以後その話者の尤度は計算する必要がなくなり、HMMのフレーム尤度計算量が削減でき、これによって高速に処理することができる。例えば、20個の話者クラスをもつ音響モデルのとき、標準話者として1個の話者クラス(話者クラスとは、複数の話者を含む1つのグループをいう。)を採用したならば尤度計算量は1/20になる。

(b) 話者適応機能を用いることにより、話者の音響モデルを入力話者に効率よく適応させるとともに、不特定話者モードでは認識しにくい話者に効果的に対処することができる。さらに、話者適応のための教師信号として認識動作完了後に選出した尤度最大の発声内容を用いることにより、「教師なし話者適応」が実現することができる。

(c) 従来の不特定話者音声認識システムは、「不特定話者音響モデル」が用いられ、多数の話者の音声データを混合して学習に用いたために、広がり過ぎた混合分布によって認識性能の劣化を有していた。これに対して、本実施例では、多数の話者の音声データを混合して学習する必要がないために、認識性能の劣化を回避することができ、これによって、高い認識性能を得ることができる。

【0036】

【発明の効果】以上詳述したように本発明によれば、不特定話者の音声を連続的に認識する不特定話者連続音声

認識方法であって、複数m人の話者に対応して複数m個の発声内容の仮説が存在し、その後各仮説は話者毎にそれぞれ時間経過につれて成長させた音素解析木を用いて、入力された1人の話者の発声内容に基づいて発声内容と話者の2方向を同時にサーチの対象としてビームサーチしながら音声認識を連続的に実行し、上記音声認識動作中に計算された尤度が所定のしきい値以上のときはこれ以降の尤度計算を行い認識候補として残す一方、所定のしきい値未満となったときは、尤度がしきい値未満となった枝に接続される1つ又は複数の枝を枝刈りしてこれ以降認識候補として残さないで尤度の計算を行わず、上記音素解析木の終端部において最大の尤度又は所定値以上の尤度を有する発声内容と話者とを同時に音声認識結果として決定する。従って、本発明は以下の特有の効果をも有する。

(a) 順位の低い仮説しかもたない話者は枝刈りされ、その時点から以後その話者の尤度は計算する必要がなくなり、HMMのフレーム尤度計算量が削減できる。これによって、高速に処理することができる。

(b) 従来の不特定話者音声認識システムは、「不特定話者音響モデル」が用いられ、多数の話者の音声データを混合して学習に用いたために、広がり過ぎた混合分布によって認識性能の劣化を有していた。これに対して、本発明では、多数の話者の音声データを混合して学習する必要がないために、認識性能の劣化を回避することができ、これによって、高い認識性能を得ることができる。

【図面の簡単な説明】

【図1】 本発明に係る一実施例である不特定話者音声認識装置のブロック図である。

【図2】 本実施例における話者パージング認識モードにおける音素解析木を示す図である。

【図3】 従来例における不特定話者認識モードにおける音素解析木を示す図である。

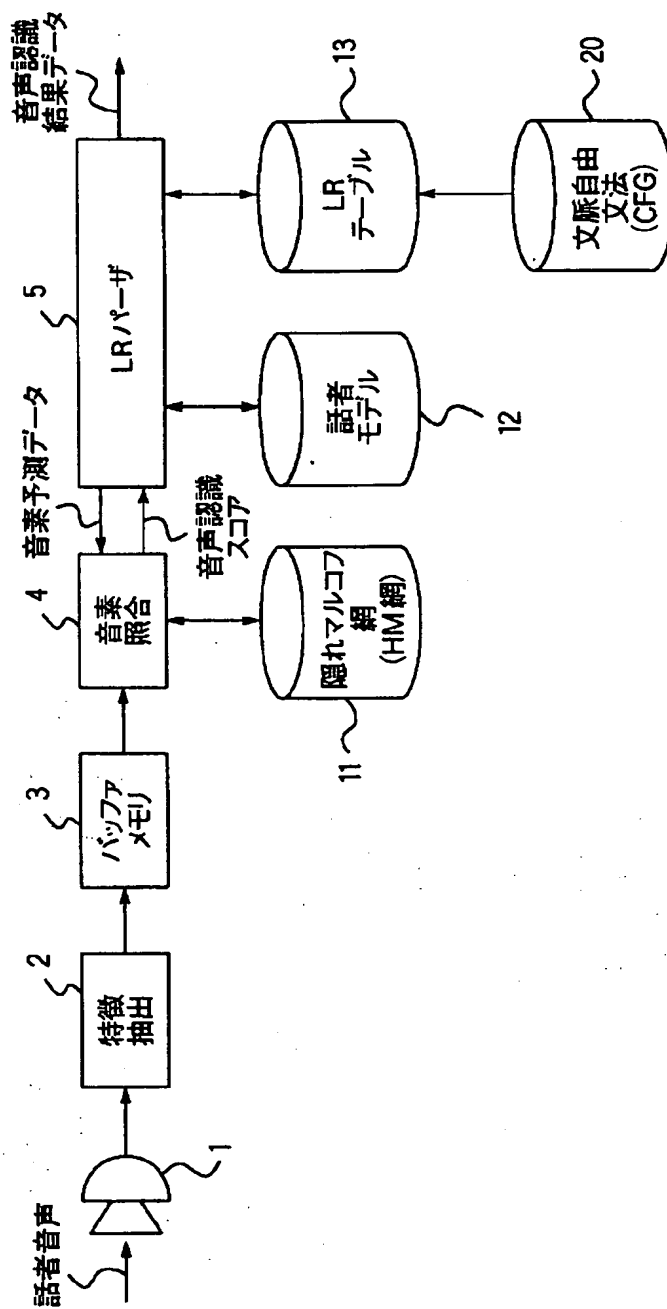
【符号の説明】

- 1…マイクロホン、
- 2…特徴抽出部、
- 3…バッファメモリ、
- 4…音素照合部、
- 5…LRパーザ、
- 11…隠れマルコフ網メモリ、
- 12…話者モデルメモリ、

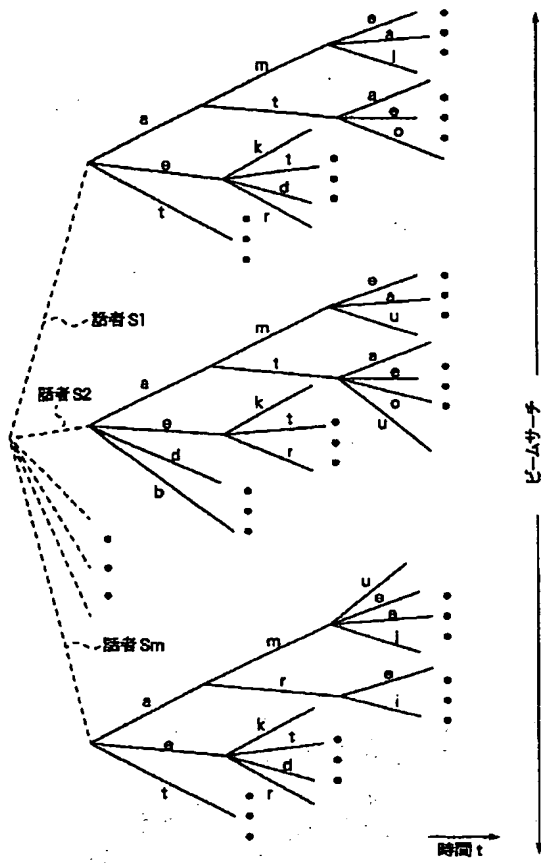
13...LRテーブルメモリ、

20...文脈自由文法データベースメモリ。

【図1】



【図2】



【図3】

